REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE
31 July 2017

2. REPORT TYPE
Technical Paper

3. DATES COVERED *(From - To)*
09 June 2017 - 31 July 2017

4. TITLE AND SUBTITLE
Optimizing Sparse Representations of Kinetic Distributions via Information Theory

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Robert Martin and Daniel Eckhardt

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
Q0A5

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory (AFMC)
AFRL/RQRS
1 Ara Drive
Edwards AFB, CA 93524-7013

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory (AFMC)
AFRL/RQR
5 Pollux Drive
Edwards AFB, CA 93524-7048

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-RQ-ED-TP-2017-171

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited. PA Clearance Number: 17446 Clearance Date: 11 July 2017

13. SUPPLEMENTARY NOTES
For submission to the IPAM Website; July 2017; (Research in Industrial Projects for Students (RIPS) program at UCLA)
The U.S. Government is joint author of the work and has the right to use, modify, reproduce, release, perform, display, or disclose the work.

14. ABSTRACT
This project is on the use of ideas from information theory in the kinetic simulation of a gas or plasma. A kinetic simulation describes the interactions (i.e., collisions and convection) of particles that constitute a gas or plasma. Since the number of physical particles is often much too large (e.g., 1020) for direct molecular dynamics computations, kinetic simulation often uses a moderate number, N (e.g., 105-107), representative "computational macro-particles" which act as surrogates for the particle interactions. The particle positions, xn, and velocities, vn, for n ranging from 1 to N, are a representative sample of a probability distribution function f(x; v). Traditionally, these macro-particles have all represented a constant number of real particles with a particle "shape" which is a single (Dirac-delta function) velocity and either delta functions in space or low order splines dependent on the spatial resolution sought as described in more detail in Bridsall’s classic reference [1]. This sparse sampling of f results in a direct trade-off between spatial accuracy and statistical noise for key flow-field parameters such as mass, momentum, energy, and physical entropy.

15. SUBJECT TERMS
N/A

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
Unclassified	Unclassified	Unclassified	SAR	7	Robert Martin
					19b. TELEPHONE NUMBER *(Include area code)* N/A

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Optimizing Sparse Representations of Kinetic Distributions via Information Theory

Robert Martin[1] and Dan Eckhardt[2]

[1]In-Space Propulsion Branch, Air Force Research Laboratory
[2]NRC Post-Doctoral Fellow

June 9, 2017

1 Introduction

This project is on the use of ideas from information theory in the kinetic simulation of a gas or plasma. A kinetic simulation describes the interactions (i.e., collisions and convection) of particles that constitute a gas or plasma. Since the number of physical particles is often much too large (e.g., 10^{20}) for direct molecular dynamics computations, kinetic simulation often uses a moderate number, N (e.g., 10^5-10^7), representative "computational macro-particles" which act as surrogates for the particle interactions. The particle positions, x_n, and velocities, v_n, for n ranging from 1 to N, are a representative sample of a probability distribution function $f(x, v)$. Traditionally, these macro-particles have all represented a constant number of real particles with a particle "shape" which is a single (Dirac-delta function) velocity and either delta functions in space or low order splines dependent on the spatial resolution sought as described in more detail in Bridsall's classic reference [1]. This sparse sampling of f results in a direct tradeoff between spatial accuracy and statistical noise for key flow-field parameters such as mass, momentum, energy, and physical entropy.

An alternative to particle samples of the probability distribution is to discretize $f(x, v)$ directly in simulations that treat the time evolution as a continuum Partial Differential Equation (PDE). This approach, referred to as Vlasov or direct kinetic methods, is free from statistical noise, but it is generally impractical in all but the lowest dimensional cases such as 1-space and 1-velocity dimension (1D1V). To illustrate this issue, consider the discretization of the full 6D (3-space, 3-velocity) distribution as a PDE. Using a relatively coarse mesh of 256 finite difference grid points in each dimension with only 1-byte of data per point, the 256^6 bytes (281 Tb) of memory exceeds the entire memory capacity of all but Thunder, the largest of the Air Force Research Laboratory (AFRL) supercomputers. This vastly exceeds the number of particles used in a similar spatial resolution for a typical particle simulation and is often referred to as the "*curse of dimensionality*". The statistical noise of the particle simulation would be mitigated significantly u sing a n e quivalent 16-million particle Degrees of Freedom (DOF) in every spatial cell even though the noise is only decreased by $N^{1/2}$. How this error compares to the discretization error of the coarse spatial mesh is needed for a true apples-to-apples comparison of the two approaches.

The research problem of this project is to explore the relationship between these sources of error and to investigate alternative shape functions and the optimal use of the numerical particle DOF

to improve the signal to noise ratio in kinetic simulations. The goal is to investigate how the continuous distribution function $f(x, v)$ can be best sampled and approximated using the discrete set of particle DOF data. If the numerical particles are allowed to vary in either shape or the number of real particles they represent, how should these "weights" vary to minimize error in the key flow-field parameters? A possible criteria for choosing what is "optimal" is to use an information theoretic quantity such as information entropy to ensure that the amount of information describing the distribution represented by each particle/DOF is maximized. This approach attempts to draw analogies between kinetic theory and optimal coding theory. Other choices for the optimality criterion could also be investigated.

2 1D Example

A useful illustration of this problem which is central to the project results from attempting to calculate the non-equilibrium kinetic physical entropy (not to be confused with the information theoretic concept of information entropy) as described in Chapter IX of Reference [2]. This quantity can be calculated as shown in Equation 1.

$$S = -\int f log(f) d\Omega \tag{1}$$

This physical entropy is maximized at equilibrium[1] when the probability distribution assumes the form of a Maxwellian.

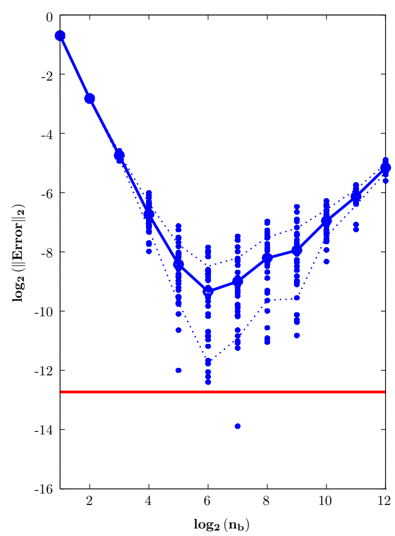

Figure 1: Tradeoff of quadrature error and statistical noise for integration of physical entropy into n_b-bins. Dots are samples of the error while lines are the mean with standard deviation bounds. Error for equal weight bins from uniform intervals of erf^{-1}(-1:1) are also included. (red).

For the purposes of this example, the problem can be further restricted to the simplest case of the spatially homogenous numerical integration of this physical entropy quantity in just one velocity dimension. The first numerical approximation to the physical entropy is then simply $\bar{f}log(\bar{f}) \cdot \Omega$ where $\bar{f}$ is simply the total number of physical particles in the domain, N, divided by some total volume, Ω, of the $(\mathbb{R}^x \times \mathbb{R}^v)$ bounding box. This provides an extremely coarse estimate for the physical entropy due to the quadrature error of the coarse bounding box. For a sufficient number of particles, this quadrature error is significantly larger than the error induced by noise in f resulting from a finite number of particles.

This estimate could be improved by progressively sub-dividing the bounding box into bins and calculating $\sum_i f^i log(f^i) d\Omega^i$ where f^i is the number of particles in each sub-bin. The refinement begins by progressively improving the estimated shape of f and therefore removes quadrature error. However, the number of particles in each $d\Omega^i$ bin simultaneous becomes smaller and smaller. This results in larger stochastic noise in the estimate for f^i which eventually dominates the quadrature error in the integration. Once this noise floor is reached, continued refinement no longer improves the estimate of the physical entropy. This effect can be seen in Figure 1 showing the error for integrating the physical entropy with 2^{16} particles randomly sampled from a unit density 1D Maxwellian

[1]This results from Boltzmann's $\mathcal{H}$-theorem in Chapter IX Section 4 [2]) and serves as the basis the second law of thermodynamics.

distribution. In the case shown, the 1D distribution was simply $f = \frac{1}{\sqrt{\pi}} e^{-\hat{v}^2}$. This assumes $\overline{v} = 0$ and the velocity has been nondimensionalized using a characteristic thermal velocity[2] such that $\hat{v} = v/v^{th}$.

If the refinement is continued anyway with uniform bins, eventually bins with zero particles per cell result. Even if these $f = 0$ cell are discarded, the volume of the cells still containing particles continues to decrease causing $f^i \to \infty$ in these cells. This makes the integral to diverge. For a given number of randomly selected particles, there therefore exists an optimal uniform bin width which extracts the most "information" about the physical entropy as is available. Below this bin width the integration is simply sampling noise. This demonstrates that the number of particles per cell and the quadrature spacing should not be considered independent parameters. This tradeoff also suggests that uniform bins in velocity space may be a poor choice of quadrature. The noise level in the tails of the distribution is much higher than the noise level near the core of the distribution for equal width bins. An optimal integration strategy would therefore attempt to balance the signal to noise ratio with the quadrature error of a cell in velocity space[3]. These issues of quadrature error and stochastic noise are in fact two sides to the same coin. Though the continuum methods are considered directly causal and immune to the stochastic errors, this is only true in the limit of infinite degrees of freedom. The same is true for the "infinite particle" limit of the stochastic methods.

For a finite dimensional continuum method, the quadrature error may be bounded by a constant that converges uniformly with number of degrees of freedom, but the actual realization of the error fluctuates depending on the location of the quadrature points with respect to the features of the flow. The finite bandwidth of the continuum representation also has the potential to produce noise via aliasing of high frequency components. A set of so-called continuum simulations of a given resolution then too can have a random distribution of error depending on the location of quadrature points just as the stochastic methods do. The convergence of the error is bounded by different parameters, but the same fundamental problem needs to be addressed. That is, *what is the most efficient distribution of degrees of freedom for accurate solutions given a finite computational cost?*

3 Analogy to Information Theory

To approach this problem, it is useful to step back and consider how best to represent the information contained in an arbitrary probability distribution in high dimensional phase space. This provides a way to define an optimal use of computational degrees of freedom. The goal of this project is to investigate the interplay between quadrature error and stochastic noise in kinetic distributions.

The equilibrium of a particle distribution can be uniquely specified with just 2 degrees of freedom assuming Galilean (translational) invariance. For a given physical domain, these degrees of freedom are the number of particles within the region and the temperature of those particles. In this context, the purpose of a kinetic simulation is to most accurately represent the evolution of the deviation from this equilibrium, ideally for a the minimum amount of computational effort. We should then strive to define the information content of this non-equilibrium δf-component of the distribution function

[2]This thermal velocity is related to the kinetic temperature and in 1D is simply the related to the second moment of probability distribution as $(v^{th})^2 = 2\overline{v^2} = \frac{2}{n} \int v^2 f dv$.

[3]Indeed in Reference [3], Ricketson attempted to improve convergence properties of the Particle-In-Cell (PIC) method by using sparse grid techniques to provide more information about the particle charge density than could be attained using a single mesh representation.

in terms of finite bandwidth "simulation channel capacity" akin to a communication channel capacity as developed in Shannon's information theory for communication in the presence of noise [4].

The deviation from equilibrium, δf, is the perturbation away from the Maxwellian distribution defined as $\delta f \equiv f - f^{eq.}$. In the literature, this is commonly thought of as the strong pointwise deviation of the from the local equilibrium Maxwellian, but it could also be alternatively defined as the the unsplit (space & velocity) deviation from an equilibrium region defined weakly with finite spatial dimension (i.e. a computational spatial cell or region). This second definition is more reasonable as physically the concept of a particle probability distribution is only meaningful in this weak sense for a discrete finite number of real particles. This means that δf can be unsplit in $(\mathbb{R}^x \times \mathbb{R}^v)$-space. The problem of efficient numerical quadrature is then transformed into seeking an efficient coding scheme for the δf-component of the arbitrary velocity distribution.

Considering how sparsely the full high dimensional probability distribution is sampled by particles, exploring how this sparse sampling relates to compressed sensing as in References [5] is expected to help provide a stronger conceptual basis for the effort. Reconciling this signal reconstruction with the idea that the most probable distribution is the distribution that maximizes the physical entropy by minimizing the additional information consistent with the samples could serve to help tie these ideas together.

Given this framework, the group should plan to explore several potential un-split phase space spanning quadratures for this δf-component of probability distribution which can minimize error for a given number of degrees of freedom. The problem can first be approached from the basis of relatively low order adaptive methods such as binary trees and then attempts to build these into up higher order methods with intrinsic conservation properties can be explored. Ideas from image and video compression may provide additional insight into efficient coding schemes.

4 Project Goals

The goal of this project is to design adaptive quadratures to efficiently represent the information content in arbitrary particle probability distributions.

1. The first step will be exploring integration quadratures that can best be used to minimize the error in integrating the physical entropy for deviations from a 1D-Maxwellian distribution. The optimal uniform quadrature as a function of the number of (uniform weight) particles used to sample the Maxwellian can first be identified. First uniform (0^{th}-order) binning can be explored, but higher order bins conserving additional moments can also be explored in terms of accuracy per degree of freedom. Then an adaptive quadratures such as binary trees can be explored to see if the error level can be reduced below this optimal uniform bin level for a given number or particles by attempting to balance quadrature and stochastic errors at every level of refinement. Can the information entropy be used as a refinement criteria? How does the quadrature that minimizes the error in the physical entropy perform for computing other moments such as density, velocity, and temperature? This approach can be further extended to other distributions such as non-isotropic Maxwellians (multi-Temperature), bi-Maxwellian (including "bump-on-tail"), and split-Maxwellians where the distribution from which the particles are sampled discretely jumps to from one distribution to another at a given velocity coordinate.
2. The next step is allowing the weight of the particles used to sample the distribution to vary. Can a function for the optimal particle weight as a function of probability density be de-

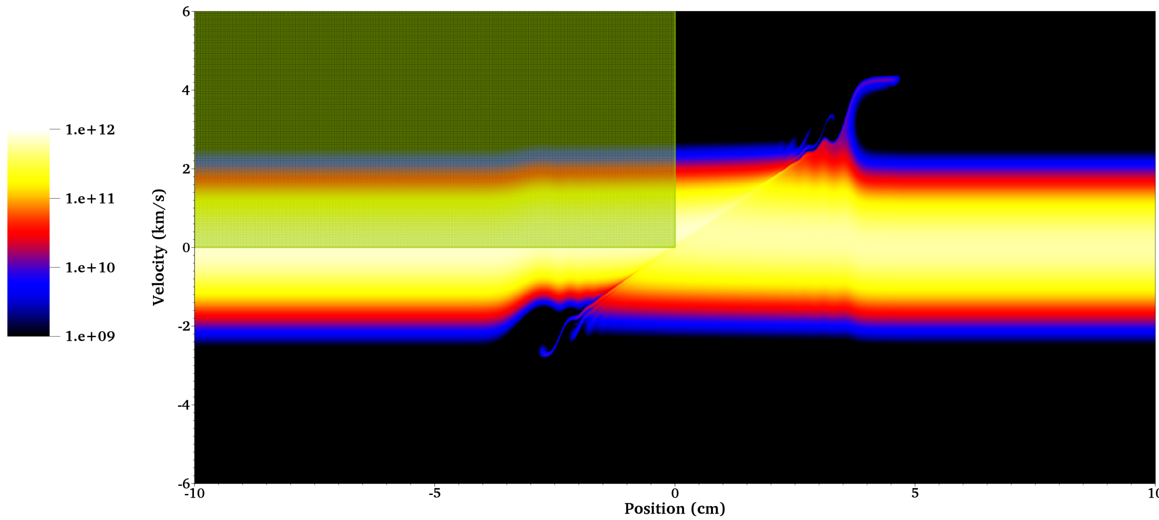

Figure 2: Vlasov probability distribution results for collisionless shock test case described in Reference [7]. Upper left corner depicts the Vlasov mesh resolution (green).

veloped? The convergence criteria for the Stochastic Weight Particle Method (SWPM) in Reference [6] may serve as a basis for this approach. Again, the goal is to minimize the error in evaluating the physical entropy for the various distributions. This can be repeated first with variable particle weights and uniform integration quadrature, and then again with adaptive quadratures. Does the adaptation criteria identified in part 1 remain the same with the variable weight particles? Is there a superior strategy for the various distributions if the adaptation criteria and particle weight strategy are simultaneously modified?

3. The next step is to subtract the sampled equilibrium distribution from the variable weight particle distribution and repeat the exploration of particle weight and quadrature adaptation strategies. How can the information content of the δf signal be quantified? Can the integration quadrature be designed to ensure that the 0^{th}, 1^{st}, and 2^{nd}-moments are identically zero so that the equilibrium distribution carries all of those pieces of information about the distribution?
4. Time permitting, the study can be extended to one velocity dimension (V) and one space (X) dimension. First with uniform weight particles, the same questions can be asked for the distributions with spatially varying (linear, sinusoidal, step) densities. Can an unsplit (XV) adaptation method out-preform tensor-product adaptation? Can the rules for particle weight and adaptation be applied in the unsplit setting or do they need to be further modified? If sufficient progress is made, the methods can also be tested by sampling particles from high resolution Vlasov results of a collisionless shock numerical experiment provided by the AFRL as described in Reference [7] and shown in Figure 4. Comparing the number of particle DOF required in the sampling to match or approach the accuracy of various levels of decimated continuum results will be of particular interest.

5 Suggested Reading

Chapter I and Chapter II Sections 1-5 of Reference [2] provide a good introduction to kinetic theory. Comparing the physical entropy of Chapter IX Section 4 of [2] with the information entropy introduced in Shannon's [4] paper will also help prepare the group for engaging in the project.

References

[1] C. K. Birdsall and A. B. Langdon, *Plasma Physics via Computer Simulation.* IOP Publishing Ltd, 1991.

[2] W. Vincenti and C. Kruger, Jr, *Introduction to Physical Gas Dynamics.* Krieger Publishing Company, 2002.

[3] L. F. Ricketson and A. J. Cerfon, "Sparse grid techniques for particle-in-cell schemes," *Plasma Physics and Controlled Fusion*, vol. 59, no. 2, p. 024002, 2017.

[4] C. E. Shannon, "Communication in the presence of noise," *Proceedings of the IRE*, vol. 37, no. 1, pp. 10 – 21, 1949.

[5] E. J. Candès, "Compressive sensing - a 25 minute tour," 2010.

[6] S. Rjasanow, T. Schreiber, and W. Wagner, "Reduction of the number of particles in the stochastic weighted particle method for the boltzmann equation," *Journal of Computational Physics*, vol. 145, no. 1, pp. 382 – 405, 1998.

[7] R. S. Martin, H. Le, D. L. Bilyeu, and S. Gildea, "Plasma model V&V of collisionless electrostatic shock," in *56th Annual Meeting of the APS Division of Plasma Physics*, (New Orleans, Louisiana), Oct 2014. Poster.

6 List of Acronyms

AFRL	Air Force Research Laboratory
DOF	Degrees of Freedom
NRC	National Research Council
PDE	Partial Differential Equation
PIC	Particle-In-Cell
SWPM	Stochastic Weight Particle Method